Machine Learning based Potato Leaves Disease Detection

# Raju Deepak Potnuru

Dept. of Information Technology

Velagapudi Ramakrishna Siddhartha Engineering College

Vijayawada, India

rajudeepak359@gmail.com

# Likhitha Portigadda

Dept. of Information Technology

Velagapudi Ramakrishna Siddhartha Engineering College

Vijayawada, India

likhitha371@gmail.com

# Dr. Sangeetha Yalamanchili

Dept. of Information Technology

Velagapudi Ramakrishna Siddhartha Engineering College

Vijayawada, India

sangeetha18.yalamanchili@vrsiddhartha.ac.in

# Tejaswini Tata

Dept. of Information Technology

Velagapudi Ramakrishna Siddhartha Engineering College

Vijayawada, India

tejaswinitata@gmail.com

# Pavan Burada

Dept. of Information Technology

Velagapudi Ramakrishna Siddhartha Engineering College

Vijayawada, India

buradapavan199@gmail.com

***Abstract*—The primary source of food, money, and employment for rural residents in economically developing nations is agriculture. Crop loss caused by plant diseases, which reduces production by 20 to 30%, is the main factor affecting agriculture productivity. Conventional methods have been used to diagnose the diseases in an attempt to prevent such losses, but they are inaccurate. To avoid losses brought on by such illnesses, accurate and prompt detection of plant diseases is vital. But occasionally those harvests and grains suffer a significant amount of damage, if not complete destruction, due to a lack of suitable cultivating knowledge, expertise, and sense of disease prediction. So, in order to lessen the loss caused by infections of plant leaves, this research attempts to integrate a portion of agriculture with the use of artificial intelligence. We used CNN transfer learning models such as VGG16, VGG19, and InceptionV3, to overcome this issue. To determine which strategy performs best at identifying potato leaf illnesses, we conducted trials using all three approaches on the standard dataset of potato leaves.**

***Keywords—Plant diseases, Convolutional neural network, Visual Geometry Group, Inception V3.***

# I. INTRODUCTION

The agriculture is by far the most prevalent occupation in the world, however there are many other types as well. India's economy heavily relies on agriculture, and this is a well-known and established fact. The key factor is expanding potato production because the demand is steadily rising worldwide and it is necessary to sell as much as our region can. Subsequent to wheat, rice, and maize, Potatoes rank as the fourth most produced agricultural food crop globally[6]. Potatoes provide large amounts of potassium, fiber, and vitamins (especially C & B6). Potatoes are a great source of soluble fiber, which can be safely consumed by individuals with high cholesterol. In addition to regulating cholesterol levels, consuming potatoes provides numerous health benefits to the body. The negative effects of diseases on crops and farming lands cannot be ignored. The underlying causes of these disorders include genetic abnormalities, viruses, and bacteria. The bulk of fungus and bacteria that make people sick are found in potato leaves. Plant diseases include fungal infections like late and early blights and bacterial afflictions like soft rot and common scab. We are motivated to create an automated method that might boost crop output, boost farmer profit, and contribute more to the economy of the country because this important plant is exposed to these diseases. On the injured potato leaf, the illness can occasionally be seen. The plant's leaves can occasionally get spots as well. Both early blight and late blight are typical illness of potatoes. The symptoms of early blight are typically characterized by small black lesions, while those of late blight can present as blister-like markings before the eventual onset of rot and drying out [7]. Farmers will benefit greatly from the deployment of the CNN algorithm to discern between these illnesses and potato leaves. So, A concept of Deep learning that uses CNN pre-trained models is shown in this study.

# II. RELATED WORK

This chapter focuses on previous research about the domain i.e., Disease prediction. It gives a brief about how the past researches and technologies were and how they are compared with the present technology we use. This is a collection of data from past.

1. In this study the researchers proposed a new approach combining SVM (Support Vector Machine) and PSO (Particle Swarm Optimization) mainly focuses on identifying the diseased portion and healthy portion of leaf. They considered four types of diseased leaves datasets as an input. The integrated concepts are implemented in MATLAB with GUI. Performance comparison among SVM, Improved SVM, and Integrated SVM with PSO are shown in terms of accuracy.
2. Study on recognition method of rice disease based on image aims to identify the three types of rice diseases by using image processing and Bayes discrimination method. Here,the disease spots are separated by outlines with different colors along with different shapes and texture features are recognized. To select effective recognition parameters the stepwise discriminant is applied. The results are calculated based on number of texture parameters.
3. In this study the researcher’s main objective is to create a standard eggplant dataset and field conditions for five major diseases. Here, the VGG16 architecture used to convert images with different variations in colors and also for feature extraction for classifying diseases using MSVM (Multi-Class Support Vector Machine).
4. In this study disease classification is done with the help of EfficientNet deep learning architecture. Here, the PlantVillage dataset is taken as a input for training the model. Results showed that B4 and B5 models of EfficientNet architecture got higher accuracies. All the models of EfficientNet architecture, the input image size was resized due to hardware limitations.
5. This study mainly focuses on classification and identification of tomato leaf diseases by considering four CNN architectures (VGG-16, VGG-19, ResNet, Inception V3) along with feature extraction and parameter-tuning process. These models did notperform well on the field-bases dataset except, Inception V3 which performed well on both field-based data, and Laboratorybased dataset.

# III. PROPOSED METHODOLOGY

1. *The Plant Village Dataset.*

There are 54,304 pictures in the PlantVillage dataset. Apple, Blueberry, Cherry, Corn, Grape, Orange, Peach, Bell Pepper, Potato, Raspberry, Soybean, Squash, Strawberry, and Tomato are among the 14 crop species represented in the photographs. It includes illustrations of 17 bacterial diseases, 2 molds (oomycete) diseases, 2 viral infections, and 1 mite disease. Images of healthy leaves from 12 crop species that are not diseased are also available [8].

Each fruit and vegetable folder has both colored and grayscale photos. For classification purposes, each type of leaf disease that affects a crop is categorized as a different class of disease. The dataset includes two pictures of leaves—one with and one without a background—one for each of the two categories.

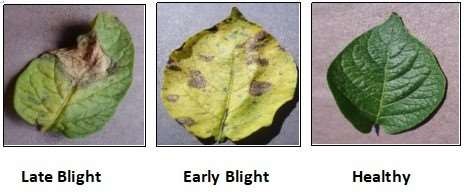


Fig.1. shows an illustration of each class.

A class may have anywhere from 152 and 1000 images, depending on the class. For our categorization study, which consists of three classes—Early Blight, Late Blight, and healthy leaf shots—we have exclusively used potato photos.

**Table 1:** Displays the different groups available in our dataset. Any transformations don’t affect this.Herewemadeuseofthreefeaturedescriptors.



1. *Image Preprocessing.*

To learn the relevant features from the image dataset preprocessing is a needed step in development of CNN models.

We have used some Image Preprocessing techniques that showed below:

**Resizing:**

We have resized the images into 224\*224 so that all the input images have set into consistent size and model performance can be improved.

**Data Augmentation:**

The Dataset is augmented by applying various transformations. And this is used to get more different image. This can also prevent overfitting to our model. In our paper, we used some of the Data Augmentation techniques. They are:

* Rescaling:

The image is rescaled by 1. /255, so that the intensity of pixel values is scaled down to the range of 0 to 1. This technique will normalize the pixel values and make the training process more stable.

* Shear range:

Shear range with a value of 0.2 is used. Means that the maximum angle of the shear transformation will be 0.2 radians (or about 11.5 degrees). This results a slanting or skewing effect.

* Zoom range:

Zoomtransformation of 20% i.e.,zoom\_range(0.2) is used. This results a zooming in effect.

* Horizontal flip:

The input images are randomly flipped horizontally to create new variations during training.

**Regularization:**

In order to avoid overfitting and guarantee that the model generalizes effectively to new data, regularization techniques like dropout, L1 or L2 regularization, and early stopping is used. In our paper we used L2 regularization with a strength of 0.01 is applied to the weights of layers during training, Dropout (0.5) is used so that half of the inputs to the layer are randomly set to zero. And Early stopping is used to find the optimal point in the training process where the model achieves best performance without overfitting the training data.

**Splitting the dataset:**

To overcome unbiased results, we have divided the preprocessed dataset into training set, testing set, and validation set.

1. *Classifiers Models.*

**CNN Pre-Trained Models:** In this paper we have used CNNpretrained models such as vgg16, vgg19, inception v3.Computer Vision Problems often don't have large datasets, making it difficult to achieve good accuracy even with data augmentation techniques. Using deep learning models with many parameters can result in overfitting. Transfer learning can help to overcome this issue. Transfer learning involves using a pre-trained model and fine-tuning it for a specific task, rather than training a model from scratch [9].

All the pretrained models we used loss function i.e.Sparse categorical cross entropy is used since our target variables in classification problem are integers rather than one-hot encoded vectors. The sparse categorical cross entropy calculates the loss based on these integer labels and is used for multi-class classification problems with a large number of classes [10].

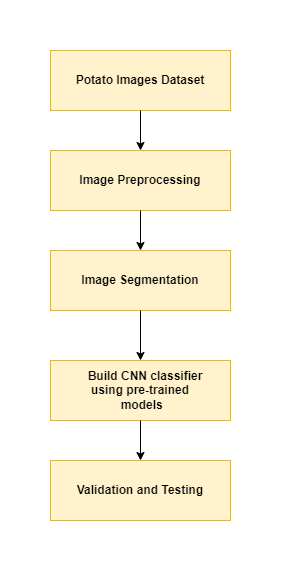


Fig.2. shows flow of methodology

1. *Performance Comparison among pre-trained networks:*

The performance of multiple pre-trained deep learning models, including VGG16, VGG19, and InceptionV3, has been assessed. In this paper, we have run Each of the three models for 20 epochs with early stopping.

**InceptionV3:**

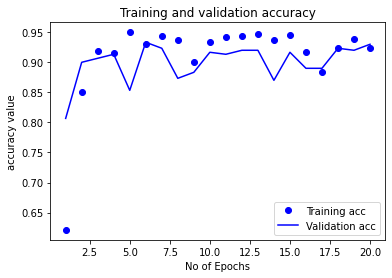
****

Fig.3. shows performance of InceptionV3

On training the dataset with InceptionV3 model, we got a higher accuracy of 94.56% at Epoch 15. This model gives an accuracy of 93.6% on test dataset. And gives an accuracy of 93% on validation dataset. So, this model is performing well on unknown dataset.

**VGG19:**

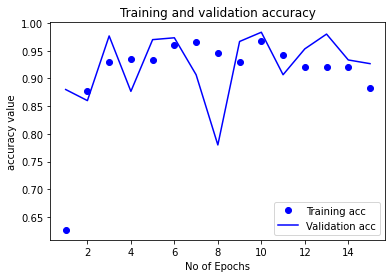
****

Fig.4. shows performance of VGG19

On training the dataset with VGG19 model, we got a higher accuracy of 96.78% at Epoch 10. This model gives an accuracy of 94.6% on test dataset. And gives an accuracy of 92.6% on validation dataset. So, this model is also performing well on unknown dataset.

**VGG16:**

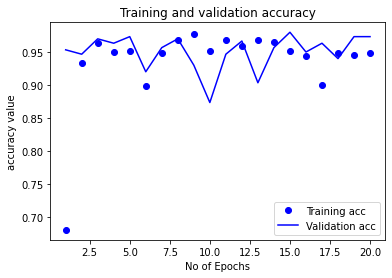
****

Fig.5.shows performance of VGG16

On training the dataset with VGG16 model, we got a highest accuracy of 97.78% among VGG19, and InceptionV3 at Epoch 9. This model gives an accuracy of 98% on test dataset. And gives an accuracy of 97.3% on validation dataset. Among VGG19, and InceptionV3 this model is giving higher accuracy, and also shows better performance in the classification of unknown data.

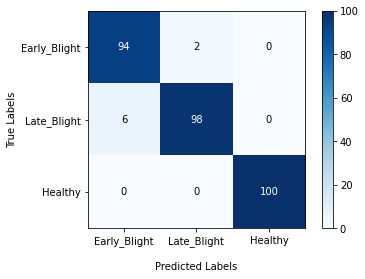


Fig.6. shows confusion matrix of VGG16 on Validation dataset

From the confusion matrix it is observed that the model is predicted Healthy leaves perfectly. Out of 100 Late Blight images the model predicted 98 images that belongs to the actual class. And in the case of Early Blight, it predicted 94 images that belongs to the actual class. VGG16 is performing well on unknown dataset.

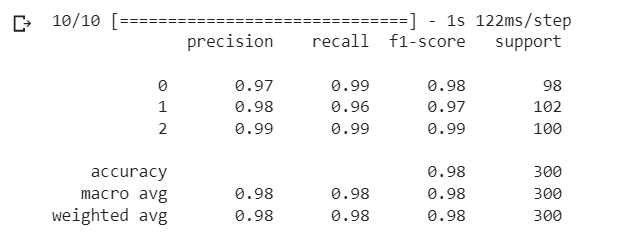


Fig.7. shows classification report of VGG16

# IV. EXPERIMENTS AND RESULTS

Finally, we have done the classification of potato leaf disease images using three pre-trained models (VGG16, VGG19, Inception V3). Among them VGG16 gave us precious results. The study was performed using Google Collaboratory, a machine learning educational and research tool offered by Google. This Jupyter notebook environment is user-friendly and requires no setup. It is provided for free and allows for the execution of codes that necessitate high-performance GPUs [10].

# IV. CONCLUSION

The pre trained models of Convolutional Neural Network (CNN) are used in this research study. The proposed model has achieved an accuracy of 97.3% in VGG16 in comparison to VGG19 that achieved 92.6 % and Inception V3 produced 93% on validation dataset.

An expanded study can be conducted on a larger dataset, incorporating a greater number of classes and more images per class, as this will likely lead to improved accuracy through increased feature learning and reduced loss. The saved weights from the model can then be utilized to develop a web or mobile application for classifying various types of leaves diseases.

# V. REFERENCES

1. Prabhjeet Kaur, Sanjay Singla, Sukhdeep Singh, Detection and classification of leaf diseases using an integrated approach of support vector machine and particle swarm optimization, International Journal of Advanced and Applied Sciences, March 2017.
2. Z.-X. Guan, J. Tang, B.-J. Yang, Y.-F. Zhou, D.-Y. Fan, and Q. Yao, ‘‘Study on recognition method of rice disease based on image,’’ Chin. J. Rice Sci., vol. 24, no. 5, pp. 497–502, May 2010.
3. K. Rangarajan and R. Purushothaman, ‘‘Disease classification in eggplant using pre-trained VGG16 and MSVM,’’ Sci. Rep., vol. 10, no. 1, pp. 1–11, Dec. 2020.
4. Atila, M. Uçar, K. Akyol, and E. Uçar, ‘‘Plant leaf disease classification using EfficientNet deep learning model,’’ Ecol. Informat., vol. 61, Mar. 2021, Art. no. 101182.
5. Ahmad, M. Hamid, S. Yousaf, S. T. Shah, and M. O. Ahmad, ‘‘Optimizing pretrained convolutional neural networks for tomato leaf disease detection,’’ Complexity, vol. 2020, pp. 1–6, Sep. 2020.
6. Kothari, Deep, Harsh Mishra, Mihir Gharat, Vishal Pandey, and Rashmi Thakur. "Potato Leaf Disease Detection using Deep Learning."
7. Asif, Md Khalid Rayhan, Md Asfaqur Rahman, and Most HasnaHena. "CNN based disease detection approach on potato leaves." In 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), pp. 428-432. IEEE, 2020.
8. Divyansh Tiwari, Mritunjay Ashish, Nitish Gangwar, Abhishek Sharma, Suhanshu Patel, Suyash Bhardwaj. "Potato Leaf Diseases Detection Using Deep Learning", 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020.
9. Ishengoma, Farian Severine, Idris A. Rai, and RutabayiroNgoga Said. "Identification of maize leaves infected by fall armyworms using UAV-based imagery and convolutional neural networks." *Computers and Electronics in Agriculture* 184 (2021): 106124.
10. J. R. Rajayogi, G. Manjunath and G. Shobha, "Indian Food Image Classification with Transfer Learning," 2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), Bengaluru, India, 2019, pp. 1-4, doi: 10.1109/CSITSS47250.2019.9031051.
11. V. Singh and A. K. Misra, ‘‘Detection of plant leaf diseases using image segmentation and soft computing techniques,’’ Inf. Process. Agricult., vol. 4, pp. 41–49, Mar. 2017.
12. A. Krizhevsky, I. Sutskever, and G. E. Hinton, ‘‘ImageNet classification with deep convolutional neural networks,’’ in Proc. Adv. Neural Inf. Process. Syst. (NIPS), vol. 25, Dec. 2012, pp. 1097–1105.
13. Y. Li, J. Nie, and X. Chao, ‘‘Do we really need deep CNN for plant diseases identification?’’ Comput. Electron. Agricult., vol. 178, Nov. 2020, Art. no. 105803
14. G. Geetharamani and P. J. Arun, ‘‘Identification of plant leaf diseases using a nine-layer deep convolutional neural network,’’ Comput. Electr. Eng., vol. 76, pp. 323–338, Jun. 2019.
15. E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, ‘‘A comparative study of fine-tuning deep learning models for plant disease identification,’’ Comput. Electron. Agric., vol. 161, pp. 272–279, Jun. 2019.